R E P O R T R E S U M E S

ED 014 375 RE 000 236

AN ANALYSIS OF EIGHT DIFFERENT READING INSTRUCTIONAL METHODS USED WITH FIRST GRADE STUDENTS.

BY- CAMPBELL, PAUL B. AND OTHERS

PUB DATE MAY 67

EDRS PRICE MF-$0.25 HC-$1.00 23P.

DESCRIPTORS- *READING MATERIALS SELECTION, *READING MATERIALS, *READING INSTRUCTION, *READING PROGRAMS, READING ACHIEVEMENT, BEGINNING READING, *READING RESEARCH, LIVONIA PUBLIC SCHOOLS, MICHIGAN,

DURING THE SCHOOL YEAR 1965-66, GROUPS OF FIRST-GRADE PUPILS IN THE LIVONIA SCHOOL SYSTEM, MICHIGAN, WERE PROVIDED WITH READING INSTRUCTION MATERIALS FROM EIGHT PUBLISHERS REPRESENTING EIGHT APPROACHES AS PART ONE OF A 2-YEAR STUDY. THE MATERIALS WERE THE LIPPINCOTT BASIC READING SERIES, THE SCIENCE RESEARCH ASSOCIATES BASIC READING SERIES, THE SCOTT FORESMAN BASAL READING SIXTIES SERIES, THE GINN BASIC READERS, THE MICHIGAN SUCCESSIVE DISCRIMINATION LANGUAGE READING PROGRAM, THE STERN STRUCTURAL READING SERIES, THE MCGRAW-HILL (SULLIVAN) PROGRAMED READING MATERIALS, AND THE INITIAL TEACHING ALPHABET PROGRAM. PRELIMINARY MEASURES WERE COLLECTED ON MENTAL ABILITY, READING READINESS, SOCIOECONOMIC STATUS, TEACHER ESTIMATE OF READING MOTIVATION, INTEREST IN READING, DOMINANCE, AND PRESENCE OR ABSENCE OF SPEECH PROBLEMS. THE FOLLOWING QUESTIONS WERE CONSIDERED IN THE COMPARISON--WHICH METHOD (1) PRODUCES THE HIGHEST LEVEL OF ACHIEVEMENT, (2) PRODUCES HIGHEST LEVEL OF WRITING ABILITY, (3) SEEMS BEST SUITED FOR CHILDREN OF HIGH OR LOW PREVIOUS INTEREST IN READING, (4) PRODUCES A SPECIAL TYPE OF READING ACHIEVEMENT, (5) IS BEST SUITED FOR HIGH OR LOW READINESS LEVELS, (6) IS BEST FOR GIRLS OR BOYS, (7) IS BEST FOR CHILDREN WITH SPEECH PROBLEMS. A DETAILED DISCUSSION OF THE PROCEDURES AND FINDINGS IS GIVEN. THIS PAPER WAS PRESENTED AT THE INTERNATIONAL READING ASSOCIATION CONFERENCE (SEATTLE, MAY 4-6, 1967). (RH)

ED014375

Paul B. Campbell, Coordinator of
Instructional Research and
Development Services
Henry C. Heusner, Director of
Elementary Education
June J. Slobodian, Coordinator of
Reading Improvement Services
Livonia Public Schools
15125 Farmington Road
Livonia, Michigan 48154

IRA

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

RE 000 236

AN ANALYSIS OF EIGHT DIFFERENT READING INSTRUCTIONAL METHODS USED WITH FIRST GRADE STUDENTS

Session XI - Research in the Pedagogy of Reading

School districts, in approaching the problem of selecting materials on instructional methods for the teaching of reading, are frequently bewildered by the varying claims of publishers and authors of such programs. It is nearly always possible to find in the literature evidence which supports the success of each program with some group of children somewhere. The question for the school district then becomes one of evaluating these claims and reported successes in terms of the particular set of conditions which operate on the local scene.

Systematic study of the output of the various programs, given the local conditions, is a task which is too often neglected. Such a

RE 000 236

study is not basic research in the traditional sense, although the replication of many local efforts across the country may well provide basic knowledge about the effectiveness of various kinds of programs. It is, nevertheless, a very important and even critical activity in which local school districts should engage.

During the 1965-66 school year, groups of first grade children in the Livonia school system were provided with reading instruction material from eight different publishers, representing eight different approaches or materials, as the first part of a two-year study. The materials were: Lippincott Basic Reading Series (1); Science Research Associates Basic Reading Series (2); Scott Foresman Basal Reading Sixties Series (3); Ginn Basic Readers (4); Michigan Successive Discrimination Language Reading Program (6); Stern Structural Reading Series (7); McGraw-Hill (Sullivan) Programed Reading Materials (8); and the Initial Teaching Alphabet (i/t/a) (9) program. Preliminary measures were collected on mental ability, reading readiness, socio-economic status, teacher estimate of reading motivation, interest in reading, dominance, and presence or absence of speech problems.

Materials were assigned by an essentially random procedure among those teachers who were willing to volunteer for new reading materials. The Ginn Basic Readers material served as a control, with an added advantage in that it, too, was a new adoption for the 1965-66 school year. Thus, all of the materials were essentially new to the people who were using them in the classroom, although, in each case, some previous familiarization had taken place before actual teaching began.

The assumption of equivalence among groups was tested by means of an analysis of variance of the pre-measures on mental ability and

reading readiness. No significant differences among groups were found on mental ability. A difference appeared, however, on reading readiness. The entire study group was, therefore, divided into subsets for the purpose of analysis.

The Lippincott and McGraw-Hill materials formed a subset with the Ginn, while the remaining materials formed another subset overlapping Ginn, but not including Lippincott and McGraw-Hill. In other words, there were no significant differences between the Ginn control group and the Lippincott and McGraw-Hill groups, nor between the Ginn control group and the remaining groups in the study. There were, however, significant differences between Lippincott, McGraw-Hill and the groups using materials such as the Michigan Successive Discrimination Language Program.

After a year's instruction, during which all teachers were provided with appropriate consultative service and demonstration help when needed, a preliminary analysis of reading achievement was made. The children were tested using the Stanford Achievement Test, Primary Battery I, 1964 edition. Prior to the analysis of the data, the differences in group size were moderated somewhat by random sampling from the population of children who were working with any given material. Test scores were tabulated and the Hartley F_{max} tests were used to assess the validity of the assumption of homogeneity of variance. This assumption was supported for the smaller experimental subset and for the larger experimental subset with one exception; a rather large divergence in the Word Study Skills section of the Stanford Achievement Test. The data were, however, studied by means of the analysis of variance model with the divergence accounted for, where possible, by

alternative analyses. Further random samples, categorized on the basis of variables of interest (i.e., male and female, high and low readiness, etc.) were also drawn, were subjected to the same tests, and were analyzed by appropriate statistical methods, including distribution free statistical tests.

Tentative Conclusions:

The comparison of the eight commercial materials and reading instruction methods was undertaken to provide a basis for answering the following questions:

1. Which among the eight methods and materials seem to produce the highest level of reading achievement for children in general?
2. Which among the eight methods and materials seem to produce the highest level of writing ability for children in general?
3. Are there reading materials and methods which seem to be uniquely suited for use in instructing children of high or low previous interest in reading?

   Corollary to this question, is it possible to collect accurate information about levels of reading interest?
4. Are there differences among the materials and methods in producing types of reading achievement?
5. Do groups of children categorized by high and low levels of readiness perform better with some materials than with others?
6. Do girls perform better than boys with certain types of materials and methods?
7. Are some materials uniquely suited for teaching children with speech problems?

8. Are some materials and methods better suited for teaching children with mixed dominance patterns?

The findings support the following conclusions:

With respect to question one, no significant F ratios were found. The data, therefore, give us reason to believe that Livonia children, in general, have no advantage or disadvantage when they are taught by Lippincott linguistic or McGraw-Hill programed reading as opposed to Ginn reading material.

In the larger experimental subset, however, the groups ranked themselves in the following order:

The i/t/a group equals or significantly exceeds the Ginn group in all five subtests.

Scott Foresman Sixties group does likewise, except that the Scott Foresman group exceeds the Ginn group in only one subtest, Vocabulary, ($t = 2.59$ df 6,120)* while the i/t/a group exceeds Ginn in three: Spelling ($t = 3.53$ df 6,120), Word Study Skills ($t = 3.86$ df 6,120), and Word Reading ($t = 5.61$ df 6, 120).

The group using the Stern material does not differ from the Ginn group in Vocabulary, Spelling, or Word Study Skills. The obtained means were lower than Ginn in Spelling and Word Study Skills, but not significantly so. The Stern material is significantly poorer on Paragraph Meaning ($t = 2.67$ df 6,120) and Word Reading ($t = 2.81$ df 6,120) than the Ginn group.

The Michigan Successive Discrimination program equals the Ginn material on Vocabulary and Word Study Skills. This group does less

*Reported t values are computed using Dunnett's method for comparing multiple means with a control (10).

well than the Ginn group on Paragraph Meaning (t = 2.67 df 6,120), Spelling (t = 3.53 df 6,120), and Word Reading (t = 2.81 df 6,120).

The SRA group equals the Ginn group in Vocabulary; it does not do as well as the Ginn group in Paragraph Meaning (t = 2.67 df 6,120), Spelling (t = 2.64 df 6,120), and Word Reading (t = 2.81 df 6,120). The Word Study Skills mean is lower than the Ginn mean, but the divergent nature of the variances makes this result uncertain. There is, however, no reason to believe that the SRA material was as successful or more successful than the Ginn material in Word Study Skills.

It would appear from these data that the i/t/a material is promising enough to be given a further tryout, and that the Ginn and Scott Foresman Sixties materials succeed quite well in Livonia classrooms. The same statement may be made for the Lippincott and McGraw-Hill materials, although the latter was judged by the staff as quite difficult to employ. The other materials do not offer much general promise in the Livonia school setting. Some cautions are, however, in order. First, previous experience with Stern material seems to indicate that the problem in Paragraph Meaning tends to disappear by the end of the second grade. Second, the i/t/a material needs to be re-evaluated at the end of the second grade likewise, since it is possible that the nature of the test, printed in i/t/a rather than traditional orthography for the i/t/a group, favored children who had been trained by the i/t/a method. There is little doubt, however, that a substantial plurality had a very successful first grade experience with i/t/a. Finally, further testing is essential because these results may be generalized only to the conditions operating this year. A follow-up of these children, and a replication of the study with the more promising materials is underway.

The second question concerns quality of written expression. On the basis of a significant difference obtained by a non-parametric analysis of variance of judgments of writing samples, it appears that writing ability is best achieved in groups using SRA, i/t/a, Lippincott, and McGraw-Hill materials, in that order. Again, follow-up is required to better assess this finding.

With respect to question three, no method of instruction seems to be uniquely appropriate for high or low interest levels. The data collected do, however, support a conclusion that a measure of interest in reading can significantly predict a subsequent level of achievement after a year's instruction ($F = 18.87$ df 1,123). While this finding answers a question posed as a corollary in this study rather than a direct concern, it does open the possibility for further exploration of methods for improving reading instruction.

Concerning question four, the data support the conclusion that the eight materials of the study do vary considerably in the kind of reading achievement which they most effectively foster ($F = 32.39$ df 4,490). Word Study Skills emerge most consistently strong, and Paragraph Meaning is the most consistently weak aspect of reading achievement. The specific relation between materials and type of achievement is discussed under question one.

The data support conclusions ($H = 254.4$) concerning question five that for Paragraph Meaning, i/t/a, Ginn, and Scott Foresman materials, in that order, produce better achievement in the low readiness groups, while no significant differences occur in the high readiness groups. The Spelling subtest seems to indicate that in the low group, the i/t/a material ($t = 3.90$ df 7,62) is associated with the highest success, while the Stern material is associated with the lowest success

(t = 5.46 df 7,62). Again, no differences appear among the high groups. The previously mentioned caution about the i/t/a version of the Stanford Achievement Tests applies here. The significant H statistics (263 and 251) for the Word Reading analysis confirm again the differences among materials for all children in the study. No precise probability statement abcut interaction between levels of readiness and materials is possible here, but the data seem to indicate that Ginn, i/t/a, and Scott Foresman materials are likely to be more productive in the low groups, and i/t/a, Scott Foresman, and McGraw-Hill materials seem to fill this specification in the high groups.

Concerning question six, no materials seem to be uniquely suited for boys or girls. In addition, an unpredicted finding appears from these data; it is that there appears to be no difference between boys and girls in reading achievement in the first grade. This is contrary to common assumptions. Furthermore, this finding was duplicated for a different sample of Livonia first graders by a previous study (5). It appears possible that teachers in Livonia in the first grade are at least taking into account some of the differences between boys and girls which effect reading instruction, or else the differences do not exist when instruction begins.

No material appeared to be uniquely better than the others for children when there were speech problems, as question seven asks, nor was any difference in general reading performance found between the speech problem group and the no speech problem group. A slight trend in the direction of lower achievement associated with speech problems is evident, but it cannot be ruled out as a chance occurrence.

Concerning dominance patterns, as posed by question eight, no material emerges as being uniquely best for children with mixed

dominance patterns, or conversely, for children with established dominance patterns. Some differences appear among the means between the groups. For example, in Paragraph Meaning, Scott Foresman, Stern and i/t/a materials seem to be most effective for the mixed dominance group, while Ginn, Scott Foresman, McGraw-Hill, and i/t/a seem to be best for the established dominance groups. Likewise, in the Word Study Skills, i/t/a, Scott Foresman, Stern and Lippincott materials appear to produce the best achievement for the mixed group, while for the established group, Scott Foresman, i/t/a, McGraw-Hill and Stern materials seem to do the best job. These occurrences, of course, may be chance orderings, and no firm conclusion should be drawn from them. Again, the lack of differences between the overall achievement of the mixed and established dominance groups, where it could be tested, seems to indicate that these factors are being taken into account fairly well by teaching methods in use in Livonia schools at the present time.

In considering the total pattern which emerges from this study, it appears that the most significant result is the existence of highly significant differences among the materials and methods for all children, as posed by the first question.

It should be further recognized that all such findings as these are probabilistic; in other words, the results may apply to most children, but there will always be exceptions who may profit by an approach which does not show as promising general results as the others which are included in this study. Teachers should continue to try differing approaches when children do not seem to fit with the pattern which works best for the total group.

References:

1. McCracken, Glen; Walcutt, Charles C.; Bond, Mary F., and Lupica, Marion R. Basic Reading Series. Philadelphia: J. P. Lippincott Company, 1964.
2. Rasmussen, Donald and Goldberg, Lynn. Basic Reading Series. Chicago: Science Research Associated Inc., 1965.
3. Robinson, Helen M.; Monroe, Marion; and Artley, A. Sterl. Basal Reading Sixties Series. Chicago: Scott Foresman and Co., 1963.
4. Russell, David H.; Gates, Doris; and Snedaker, Mabel. Ginn Basic Readers. New York: Ginn and Company, 1964.
5. Slobodian, June J. "An Analysis of Certain Dimensions of Teacher Behavior During Reading Instruction in the First Grade." Doctoral dissertation presented at Kent State University Graduate School, March, 1966.
6. Smith, Donald E. P. and Kelingos, Judith M. Michigan Successive Discrimination Language Reading Program. Ann Arbor, Michigan: Ann Arbor Publishers, 1965.
7. Stern, Catherine; Stern, Margaret B.; Gould, Tone S.; and Gartler, Marion. Stern Structural Reading Series. Syracuse, New York: L. W. Singer Company, Inc., 1963.
8. Sullivan, M. W. and Buchanan, Cynthia Dee. (Sullivan) Programed Reading Materials. New York: McGraw-Hill Book Company, 1963.
9. Tanyzer, Harold J. and Mazurkiewicz, Albert J. Initial Teaching Alphabet. New York: Pitman Publishing Corp., Initial Teaching Alphabet Publication, Inc., 1963.
10. Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1962, p 89.

APPENDIX A

Analysis of Reading Achievement Among Different Materials and Instructional Approache

The null hypothesis for this part of the study is that there are no significant differences among the means in reading achievement for children who have been taught using the eight different materials.

The results for the Ginn, Lippincott and McGraw-Hill subset are shown below in Table I.

TABLE I

Means and F Ratios for Subset A

	Ginn	Lippincott	McGraw-Hill	F
Paragraph Meaning	20	19	19	.43
Vocabulary	22	22	26	1.75
Spelling	21	21	18	1.60
Word Study Skills	25	26	25	.12
Word Reading	20	20	20	.11
Metro. Reading Readiness	67	70	70	.99

$F_{5,95} = 3.06$

Because no significant differences were found, no further analysis of this subset was attempted.

The Ginn, Michigan Successive Discrimination, SRA, Scott Foresman 60s, and Stern materials, however, present a different picture. Table II shows the results of this analysis. Following the establishment of a significant F ratio for each of the five subtests on the Stanford Achievement Test, a comparison was run between the Ginn control group and each of the other material groups. The method used was that of a specialized t test using the t distribution published by Dunnett (Winer, 1962, p 89). Because the sizes of the groups differed, a harmonic mean of the groups was used with the Mean Square$_{error}$ in the formula for the t ratio.

On the Paragraph Meaning subtest, the i/t/a group and the Scott Foresman 60s do not differ significantly from the Ginn control group. The i/t/a mean is one point higher and the Scott Foresman 60s mean is one point lower. The Michigan Successive Discrimination group, the SRA Linguistic group, and the Stern group are, however, significantly different from the Ginn group, and in a negative direction.

On the Vocabulary subtest, the Scott Foresman 60s group differs from the Ginn group significantly, and in a positive direction, i/t/a, Michigan Successive Discrimination, and Stern all equal the Ginn group, while the SRA group shows a non-significant positive difference of one point.

On the Spelling subtest, the i/t/a group differs significantly from the Ginn group in a positive direction. The Scott Foresman 60s group also differs in a positive direction, but not significantly. The Michigan Successive Discrimination group, the SRA group and the Stern group differ in a negative direction, with the first two being significant differences.

The Word Study Skills subtest of the Stanford Achievement Battery shows these results: The i/t/a group differs significantly in a positive direction from the Ginn group. The Scott Foresman 60s group differs in a positive direction, but not

TABLE II

Comparison of Six Instructional Approaches and Materials in First Grade Reading Study

Study	Paragraph Meaning Mean	$\bar{X}_{ex} - \bar{X}_{con}$	t_D	Vocabulary Mean	$\bar{X}_{ex} - \bar{X}_{con}$	t_D	Spelling Mean	$\bar{X}_{ex} - \bar{X}_{con}$	t_D
Ginn (control	20			22			21		
i/t/a	21	1	.89	22	--	--	25	4	3.53*
MSDLP	17	-3	2.67*	22	--	--	17	-4	3.53*
SRA	17	-3	2.67*	23	1	.65	18	-3	2.64*
S-F 60s	19	-1	.89	26	4	2.59*	23	2	1.76
Stern	17	-3	2.67*	22	--	--	19	-2	1.76
	F = 3.57*			F = 2.70*			F = 15.06*		
	MS_{error} = 32.19			MS_{error} = 60.63			MS_{error} = 32.58		

Study	Word Study Skills Mean	$\bar{X}_{ex} - \bar{X}_{con}$	t_D	Word Reading Mean	$\bar{X}_{ex} - \bar{X}_{con}$	t_D
Ginn (control)	25			20		
i/t/a	34	9	3.86*	26	6	5.61*
MSDLP	23	-2	.86	17	-3	2.81*
SRA	19	-6	2.58*	17	-3	2.81*
S-F 60s	29	4	1.72	20	--	--
Stern	23	-2	.86	17	-3	2.81*
	F = 13.2*			F = 28.09*		
	MS_{error} = 137.83			MS_{error} = 29.05		

1. $F_{1,\ 99} = 3.11$ d.f. 5, 363
2. $t_{d_{5,95}} = 2.26$ d.f. 6, 120
3. *significant at or above the 5% level
4. t_D = value computed by Dunnett's method
5. $\bar{n} = 50.8$
6. N = 369

significantly. The Michigan Successive Discrimination group, the SRA Linquistic group, and the Stern Structural group all differ in a negative direction. The SRA group shows a significant difference. This subtest, however, is the one showing a great divergence in variance, and the SRA group has the most divergent variance of any of the groups in the study. To test the effects of this divergence, the SRA group was removed and the analysis was rerun. The i/t/a group remains significantly different in a positive direction, but the significance of the difference of the SRA group is unresolved.

The Word Reading subtest shows the following results: The i/t/a group differs significantly from the Ginn group in a positive direction. The Scott Foresman 60s group is equal to the Ginn group. The Michigan Successive Discrimination group, the SRA, and the Stern all differ significantly in a negative direction.

TABLE III

FIRST GRADE READING STUDY MEANS

	1 Ability score	2 Paragraph Meaning	3 Vocabulary	4 Spelling	5 Word Study Skills	6 Word Reading	7 C.A.	8 Metropolitan Rdg. Readiness	9 Socio-Economic Status
	Standardized score	Grade equivalent score	Grade equivalent score	Grade equivalent score	Grade equivalent score	Grade equivalent score	Months	Raw score	Average Rank Order
Ginn	93	20	22	21	25	20	76	67	58
i/t/a	94	21	22	25	34	26	76	62	51
Lippincott	96	19	22	21	26	20	76	70	63
McGraw-Hill	95	19	26	18	25	20	77	70	53
M.S.D.L.P.	91	17	22	17	23	17	76	62	38
S.R.A.	92	17	23	18	19	17	75	67	73
S-F 60's	93	19	26	23	29	20	76	68	52
Stern	94	17	22	19	17	17	76	65	55

Analysis of Writing Quality Among First Grade Reading Groups

A further analysis of the differences in writing quality among first grade reading groups was made. Results are shown in Table IV. For this purpose, a random sample was drawn from a standardized writing assignment by each member of the eight different groups. The writing samples were rated by two judges as to quality, and rank ordered from one to 125. A non-parametric one-way analysis of variance was applied to these data. The results show a significant difference among the groups on the quality of composition. Inspection of the average rank orders shows the i/t/a, Lippincott, and McGraw-Hill groups forming one cluster; the SRA material stands alone as the single best scorer; and the remaining materials cluster at the opposite end of the range of means. It is probable that the SRA material contributes most to the variance reported in the significant F ratio. The performance of the Lippincott and McGraw-Hill groups may be accounted for, in part, because these groups were those highest in readiness at the beginning. The i/t/a material is surprising, because this group was one of the two lowest in readiness at the beginning of the study. A follow-up at the end of the second grade should further clarify the status of writing ability among the several reading instruction groups.

TABLE IV

Analysis of Difference in Writing Quality Among First Grade Reading Study Materials (Lowest mean represents highest quality)

	Ginn	i/t/a	Lippin-cott	McGraw-Hill	MSDLP	SRA	Scott Foresman	Stern
N =	14	17	12	8	8	10	15	33
$\bar{X}$ =	67.8	51.2	46.6	44.4	93.9	37.9	62.5	76.7

$H = 21.12$

$H_{(1,99)} = 20.09$

Reading Interest Analysis of First Grade Systemwide Reading Study

The effects of interaction between differing levels of reading interest and the several methods of reading instruction were studied by means of an interview schedule which each teacher completed individually with each student. The reading interest interview schedule contained a series of questions and probes designed to elicit from the child his perception of the level of reading interest in his earlier experience and his own reactions to opportunities to enjoy books. On the basis of this interview, the students were divided into high interest and low interest groups. These two groups were composed of random samples of ten students from each method at each level. The results of the criterion test, the Stanford Achievement Test, for these groups were analyzed in a factorial design having two levels of interest, seven levels of method, and five levels of achievement type. The MSDLP method was not included in this analysis because a sufficient sample was not available from this group. The results of the analysis are shown in Table V (a) and (b).

The critical values for the F ratios shown in Table V (a), column 6 are not those which would normally be associated with a design of this sort. The reason for this is that, upon examination of the data, the assumptions of homogeneity of variance and covariance appeared to be doubtful. Therefore, the conservative critical values of the F ratio suggested by Box, and by Greenhouse and Geisser, (Winer, pages 96 and 305), were used in evaluating the hypotheses of no differences among the groups.

As indicated on the table, two significant main effects and one interaction emerged. They are factors B and C and interaction AB. Interpretation of these findings can best be considered in the light of the purposes for which the design was originally set up. The purposes are reflected in the form of the following questions:

1. Are there differences in achievement among the several methods of instruction?

   The data from this particular sample would answer this question in the negative. It should be noted, however, that for the larger sample previously analyzed in this study, significant differences among materials were found.

2. Is there a difference in achievement between high and low interest levels?

   This question is answered affirmatively by the data. The factors measured by the reading interest schedule apparently make a difference in the outcome of reading achievement.

3. Is there an interaction between methods of instruction and interest levels?

   For this sample, the answer is again negative.

4. Is there a difference in types of achievement among the students? Or, put another way, are students better in some types of achievement than in others, as measured by the Stanford Achievement Test?

   The answer to this question is affirmative. There is a highly significant difference (see Table V (a)).

5. Is there interaction between the methods and types of achievement?

   Again, this question is answered positively for the sample under study. In other words, this means that some types of achievement are more likely to occur with some methods than others.

6. Is there any interaction between interest level and type of achievement?

   The data do not show such an interaction.

7. And, finally, there is no apparent multiple interaction between interest levels, methods, and types of achievement for the sample in question.

An examination of the means reported in Table V (b) will show the probable location of some of the significant relationships uncovered by this analysis.

With respect to the significant difference for levels of interest, it may be noted from Table V (b), that of the 35 pairs of means considered across the high interest and low interest groups, only six show differences in favor of the low interest group. Of the six, four are found in the i/t/a group, and two in the Scott Foresman group. This suggests an interaction between method and level, but due to the problems with this sample, requiring very conservative F tests, a chance occurrence cannot be ruled out.

With respect to the significant F ratio relating to type of achievement, it may be noted from Table V (b), that the means range in size in such a manner that the differences within groups may be as little as four points, and as much as 14 points in the high interest category, and from a difference of three points to a difference of 14 points in the low interest category. In other words, this means that one should expect a considerable diversity in types of achievement regardless of interest level. Significant method by type interaction indicates that some methods produce a greater diversity in types of achievement than others. The means of Table V (b) indicate that i/t/a might be one of these with a range of 14 for the high group, and 12 for the low group; that the Stern method might be another in the low group with a range of 14; and that Ginn, Lippincott, and McGraw-Hill materials in the high group, likewise, might be expected to produce large differences in types of achievement.

Several interesting possibilities emerge from this analysis. First, it appears likely that the interview schedule used to determine reading interest does in fact measure a factor which can significantly predict level of achievement for most methods of instruction utilized in this study. Second, the several materials or methods seemed to differentiate significantly among the types of reading achievement, with Word Study Skills emerging as the most consistently strong factor, and Paragraph Meaning as the most consistently weak aspect of reading achievement. Finally, the problems encountered in analysing this sample, having to do primarily with lack of homogeneity of variance, are such that further experimentation is not only indicated for the purpose of checking these results, but might hold promise for other significant findings which are not supported by these data.

TABLE V

Analysis of Variances

(a) Achievement by Interest, Method, and Type

Source of Variance	SS	df	MS	F	$F_{5, 95}$
Between subjects	29,965.24	136	221.14		
A	2,999.78	6	499.96	2.93	5.12
C	3,218.08	1	3,218.08	18.87*	5.12
AC	2,865.29	6	477.54	2.80	5.12
Subjects within groups [error (between)]	20,882.09	123	170.46		
Within subjects	20,035.80	546	36.69		
B	3,398.38	4	849.59	32.39*	3.92
AB	2,862.36	24	119.26	4.54*	
BC	157.44	4	39.36	1.50	3.92
ABC	764.63	24	31.77	1.21	2.17
B x subjects within groups [error (between)]	12,852.99	490	26.23		

A = Methods
B = Achievement types
C = Interest level

*Significant at or above the 5% level.

(b) Mean Achievement by Interest, Method, and Type
(Grade Equivalent Scores)

	high Interest						Low Interest					
	Paragraph Meaning	Vocal-ulary	Spelling	Word Study	Word Reading	Range	Paragraph Meaning	Vocal-ulary	Spelling	Word Study	Word Reading	Range
Ginn	22	24	22	24	20	4	18	21	19	21	18	3
i/t/a	20	20	26	34	26	14	22	21	23	33	26	12
Lippincott	20	28	24	32	23	12	16	18	17	18	16	2
McGraw-Hill	23	35	23	30	25	12	16	21	14	19	17	5
S.R.A.	18	23	20	20	18	5	16	20	18	19	16	4
Scott-Foresman	20	24	24	28	20	8	15	20	16	18	15	5
Stern	23	22	22	28	21	7	20	24	22	33	19	14

Analysis of Reading Achievement Categorized by Readiness Levels and by Material

The comparison among groups which is of interest in this part of the study may best be expressed by the following question: Do groups of children categorized by levels of readiness perform better with some materials than with others? In order to answer this question, the data should be analyzed by means of a three-factor analysis of variance. Unfortunately, this statistical tool requires certain assumptions about the data which could be only partly met by the scores which were obtained in this study. Therefore, several less precise alternative methods had to be utilized. Specifically, the assumption of homogeneity of covariance was avoided by using a two-factor analysis of variance and treating each subtest of the Stanford series separately. The assumption of homogeneity of variance was accounted for by two alternatives. One, if only one or two groups were significantly deviant from the remainder of the groups, in terms of this assumption, these were dropped from the analysis. Two, if a more general problem seemed present, a non-parametric analysis, not requiring homogeneity of variance, was applied. The disadvantage of these procedures is that it is not always possible to directly answer the question of comparison which is posed. Given the conditions of each set of test scores, however, the most precise method which could be justified was used. The resulting inferences reflect these conditions.

The results of the analysis of reading achievement by readiness levels and by achievement are shown in Table VI (a). It appears from these data that children classified as low in readiness will perform better in Paragraph Meaning with i/t/a, Ginn, and Scott Foresman materials, in that order. The other materials tested all produced lower average achievement on this subtest. This conclusion is supported because no significant differences were found among the means of the high group, while highly significant differences appear among the means of the low group. Because separate analyses of the groups were required, a precise probability statement of the interaction between materials and level is not possible with this subtest.

Concerning the Vocabulary subtest, no differences appear in answer to the major question. The two significant results shown merely confirm the results of the preliminary analysis on differences among materials for all children, and the expected confirmation of the ability of readiness tests to predict how well children will achieve after a year's instruction. These two conclusions are supported by the significant F ratios for materials and categories respectively. Significant differences were found in the low group between the Ginn material and Lippincott material, and between the Ginn material and Scott Foresman material. Ginn was considered as a control, and the direction of the differences was positive for Scott Foresman and negative for Lippincott. In other words, the Lippincott material did less well than the Ginn, whereas the Scott Foresman group did better (see Table VI (b)).

All three F ratios for the Spelling subtest are significant. These findings may be interpreted as follows: With respect to the significant category F ratio, the data again confirm the ability of the readiness tests to predict achievement. The significant materials F ratio, however, must be interpreted in light of the significant interaction F ratio. In other words, there are certain groups at certain levels which perform better with certain of the materials. Which materials fit with which groups is shown by Table VI (b). It may be observed from this table that the significant differences all occur in the low group, and that the i/t/a material is associated with the highest success, while the Stern material is associated with the lowest success. No significant differences appear among the high groups. In reading this table, note that all comparisons are made with the Ginn Basic Reading Program, and also that the previously mentioned cautions about the i/t/a version of the Stanford Achievement Tests apply here.

The results of the analysis of the Word Study Skills subtest show no evidence of interaction between materials and levels. The significant results for both the high and low groups merely confirm again the existence of significant differences for all children among the eight materials studied.

The significant H's for the Word Reading analysis confirm again the differences among materials on this subtest for all children in the study. No precise probability statement about interaction is possible here because the non-parametric analysis of variance does not provide this information. Inspection of the average rank orders however, indicates that while Ginn, i/t/a, and Scott Foresman materials are likely to be most productive in the low group, i/t/a, Scott Foresman, and McGraw-Hill materials seem to fill this specification in the high group.

TABLE VI

(a) Reading Achievement Categorized by High and Low Reading Readiness Groups and Materi

	Paragraph Meaning		Vocabulary		Spelling		Word Study Skills		Word Reading	
	Low	High	Low	High	Low	High	Low	High	Low	High
inn (b_1)	52.75	22	20	29	19	24	50.0	30	51.8	36.65
/t/a (b_2)	63.7	28	16	25	24	29	60.4	44	67.85	65.40
ipp. (b_3)	31.75	21	15	24	17	25	32.0	31	35.25	40.40
G-H (b_4)	35.2	21	**	**	**	**	33.2	30	28.7	44.2
DLP (b_5)	30.6	21	17	28	14	20	30.55	28	25.60	34.4
A (b_6)	18.0	17	20	24	15	19	20.8	18	17.4	16.35
-F (b_7)	44.4	25	26	30	19	27	54.65	43	49.0	56.56
ern (b_8)	29.65	24	17	24	12	25	24.0	31	38.16	36.55
ategories			42.44*		81.76*					
terials	H=254.40*	F=2.13	3.63*		11.42*		H=255.13*	F=4.90*	H=263.40*	250.5
nteraction			.77		2.22*					
			F Ratio		F Ratio				$H_{.001}$ = 24.32	

*Significant at or above the 5% level. **Not included in this analysis.
Mean grade equivalent score is recorded when F ratio is given, and average rank order is recorded when H is given.

(b) Individual Comparison Among Materials by Readiness Levels

	Vocabulary $t_{Dunnett}$		Spelling $t_{Dunnett}$		
Study	Low	High	Low	High	
$b_1 - b_2$	1.97	1.05	-3.90*	-2.22	
$b_1 - b_3$	2.46*	1.32	1.56	- .44	
$b_1 - b_5$	1.47	.26	3.90*	1.77	d.f. 62
$b_1 - b_6$	0	1.32	3.12*	2.22	$t_{.05}$ = 2.35
$b_1 - b_7$	-2.95*	- .26	0	-1.33	
$b_1 - b_8$	1.47	1.32	5.46*	- .44	

Analysis of Reading Achievement Categorized by Sex and by Material

A further analysis of the Stanford Achievement scores for students participating in the first grade reading study was made by classifying scores according to the sex of the student, and according to the material with which they were taught. Again, the comparison of interest was one of interaction between materials and the other classifying variable, sex. In other words, do boys perform better with certain materials than girls. In three of the subtests, at least one but not more than two of the material groupings had to be dropped because the data from them did not fit the requirements of this type of analysis. The results of the remaining groups are shown in Table VII. The data in this table show no significant interactions. It does not appear to make any difference whether the sex of the student is male or female in determining the probable level of success with these reading materials. The significant F ratios for materials merely confirm again the fact that for all students in the study, there are differences among the materials in reading achievement. The nonsignificant categories F ratios are especially interesting because they are contrary to the usually expected findings for boys and girls. It is commonly assumed that girls learn to read more easily in the first grade than boys do. Such a finding does not emerge from the data of this study. It is also interesting to note that a previous study in Livonia, conducted with a different sample of first graders, reports the same finding.[1] While this situation may be accounted for in several ways, it appears possible that teachers in Livonia in the first grade are at least taking into account some of the differences between boys and girls which effect reading instruction, and somehow counteracting them. An alternative possibility is supported by the fact that a preliminary analysis of the readiness scores, subdivided on the basis of achievement and material, shows no signficant differences. In other words, it seems that the boys and girls in this study got about the same start.

TABLE VII

Reading Achievement Categorized by Sex and by Material

	Paragraph Meaning		Vocabulary		Spelling		Word Study Skills		Word Reading	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Ginn (b_1)	20	21	20	24	21	22	26	27	20	20
i/t/a (b_2)	18	22	21	19	24	25	36	38	26	28
Lipp. (b_3)	**	**	19	22	19	25	19	33	18	23
McG-H (b_4)	21	18	27	24	22	17	30	23	23	19
MSDLP (b_5)	16	18	21	21	14	18	20	24	16	17
SRA (b_6)	16	19	25	21	17	20	**	**	15	18
S-F (b_7)	19	20	25	22	25	24	32	29	21	20
Stern (b_8)	**	**	24	20	17	21	19	25	**	**
Categories	1.78		.917		2.61		1.31		.75	
Materials	1.73		1.129		4.41*		3.38*		7.26*	
Interaction	1.06		1.498		1.48		1.38		1.24	
	F Ratio		F Ratio		F Ratio		F Ratio		F Ratio	

*Significant at or above the 5% level. **Not included in the analysis.

[1] Slobodian, June J. "An Analysis of Certain Dimensions of Teacher Behavior During Reading Instruction in the First Grade." Doctoral dissertation presented at Kent State University Graduate School, March, 1966.

Analysis of Reading Achievement Categorized by Speech Problems and by Material

This analysis was undertaken to determine if any of the materials tried out in the first grade reading study seem to produce better achievement for children with speech problems. If some one or more of the materials was uniquely suited to such children, then interaction should appear between the categories of speech problem and no speech problem, and among the materials. Again, in this study, a preliminary analysis of levels of readiness shows no differences among the groups before instruction began.

Table VIII shows the result of this analysis. There are no significant interaction F ratios. Therefore, no material appeared to be uniquely better than the others for children with or without speech problems. The significant F ratios for materials reflect, of course, the expected difference among all children for eight different approaches to reading instruction. These differences are expected as a result of the preliminary analysis of the total groups where such a finding was supported. The lack of significant results between categories of speech problems and no speech problems is interesting, because one would suspect that the existence of speech problems might be considered evidence of a relatively low level of development of reading readiness skills, and would, therefore, expect lower scores for children presenting speech problems. There is a slight trend in the direction of lower scores in the presence of speech problems, but it is neither consistent for all subtests and materials, nor great enough to be attributed to non-chance factors. Since precision of measurement is always a problem in studies of this kind, it would not be wise to conclude that speech difficulties create no problems in reading achievement. Further study of this phenomenon should be undertaken.

TABLE VIII

Reading Achievement Categorized by Speech and Material

	Paragraph Meaning		Vocabulary		Spelling		Word Study Skills		Word Reading	
	No sp. Problem	Speech Problem	No sp. Problem	Speech Problem	No sp. Problem	Speech Problem	No sp. Problem	Speech Problem	No sp. Problem	Spee Probl
inn (b_1)	20	19	24	17	23	20	28	26	21	19
-/t/a (b_2)	22	18	22	17	25	25	31	30	26	26
ipp. (b_3)	29	17	23	21	24	22	27	28	22	20
cG-H (b_4)	20	18	26	25	19	18	27	23	21	19
SDLP (b_5)	**	**	22	23	17	16	25	19	18	15
RA (b_6)	**	**	22	21	**	**	18	18	15	16
-F (b_7)	18	18	28	26	23	22	26	29	20	18
tern (b_8)	**	**	20	21	16	20	19	25	15	18
ategories	3.13		1.93		.25		.03		.95	
aterials	.39		1.63		4.23*		2.31*		7.00*	
nteraction	.39		.46		.54		.51		.64	

*Significant at or above the 5% level.
**Not included in the analysis.

Analysis of Reading Achievement Categorized by Dominance Pattern and by Material

The establishment of patterns of dominance in children is both a function of development and of relatively permanent individual differences. That characteristics of dominance can influence ability to learn to read has been established by previous research. The purpose of this analysis was, again, to discover whether some one or more of the materials was uniquely better suited to a pattern of mixed dominance or a pattern of established dominance. The existence of a significant interaction should indicate the effect of such a material. Mixed dominance was considered to exist when a child was inconsistent in his use of his hands, eyes, legs, or other parts of the body in performing physical tasks. For example, a child might write with his right hand, but use his left eye to look through a telescope. Because equality of development was apparent among the groups, as evidenced by no significant differences in reading readiness scores, the children with mixed dominance were more likely those who had a more or less permanent mixed dominance pattern. These were rather few in number. It was, therefore, not possible to find substantial samples of children with mixed dominance in all of the material groupings. This condition required the use, in some cases, of a non-parametric analysis which does not produce a precise estimate of interaction. Table IX shows the results of this analysis.

As expected, the differences among materials for all children were, in general, supported in each subtest except the Vocabulary section. Sample size may easily account for the lack of significance here. The Vocabulary, Spelling, and Word Reading subtests could be precisely analyzed for interaction, and no significant results emerged. The existence of interaction between materials and type of dominance might be inferred from an examination of the rank orders of the averages for each material. For example, in the Paragraph Meaning subtest, Scott Foresman, Stern, and i/t/a materials seem to be most effective for the mixed dominance group, while Ginn, Scott Foresman, McGraw-Hill, and i/t/a seem most effective for the established dominance groups. The Word Study Skills subtest, like the Paragraph Meaning subtest, had to be treated by non-parametric methods within the categories. Again, the examination of the rank orders of the material averages shows a different ordering of these averages. For the mixed group, i/t/a, Scott Foresman, Stern, and Lippincott materials appear to produce the best achievement. For the established group, Scott Foresman, i/t/a, McGraw-Hill, and Stern materials seem to do likewise.

This is a very risky kind of interpretation, however, and should be checked by further study.

TABLE IX

Reading Achievement Categorized by Dominance and by Material

	Paragraph Meaning		Vocabulary		Spelling		Word Study Skilly		Word Reading	
	Mixed	Established	Mixed	Established	Mixed	Established	Mixed	Established	Mixed	Established
Ginn (b_1)	36.4	47.4	23	22	20	21	20	39.95	19	20
i/t/a (b_2)	40.15	45.95	20	20	25	24	34	51.6	26	25
Lipp. (b_3)	37.5	38.4	16	25	23	22	30	37.25	**	**
McG-H (b_4)	28.75	46.2	24	26	15	20	20	44.05	18	22
MSDLP (b_5)	25.2	34.0	25	22	16	18	24	30.4	17	17
SRA (b_6)	24.9	32.35	21	24	17	20	19	25.9	17	18
S-F (b_7)	44.8	46.45	28	26	25	25	34	51.8	21	20
Stern (b_8)	42.15	33.25	21	24	22	20	31	43.05	20	17
Categories			.79		.62					
Materials	H=219.99*		248.06*		1.08		3.80*		F=12.66* H=253.33*	F=4.96*
Interaction			.76		.56					.74
			F Ratio		F Ratio					F Ratio

$H_{.001} = 24.32$

*Significant at or above the 5% level.

**Not included in the analysis.

Mean grade equivalent score is recoded when F ratio is given, and average rank order is recorded when H is given.

Off/Research
PBC:ms
11/66